Explaining Inconsistencies in the Education Distributions of Ten Cross-National Surveys - the Role of Methodological Survey Characteristics

Ortmanns, Verena

sciendo

Journal of Official Statistics, Vol. 36, No. 2, 2020, pp. 379–409, http://dx.doi.org/10.2478/JOS-2020-0020

Explaining Inconsistencies in the Education Distributions of Ten Cross-National Surveys – the Role of Methodological Survey Characteristics

Verena Ortmanns[1]

Surveys measuring the same concept using the same measure on the same population at the same point in time should result in highly similar results. If this is not the case, this is a strong sign of lacking reliability, resulting in non-comparable data across surveys. Looking at the education variable, previous research has identified inconsistencies in the distributions of harmonised education variables, using the International Standard Classification of Education (ISCED), across surveys within the same countries and years. These inconsistencies are commonly explained by differences in the measurement, especially in the response categories of the education question, and in the harmonisation when classifying country-specific education categories into ISCED. However, other methodological characteristics of surveys, which we regard as 'containers' for several characteristics, may also contribute to this finding. We compare the education distributions of nine cross-national surveys with the European Union Labour Force Survey (EU-LFS), which is used as benchmark. This study analyses 15 survey characteristics to better explain the inconsistencies. The results confirm a predominant effect of the measurement instrument and harmonisation. Different sampling designs also explain inconsistencies, but to a lesser degree. Finally, we discuss the results and limitations of the study and provide ideas for improving data comparability.

Key words: Comparative research; cross-national surveys; survey characteristics; education.

1. Introduction

Education is a key socio-demographic variable that is measured in nearly every survey (Smith 1995). Education is central in social stratification research, for instance, when analysing educational inequalities and how social class origin affects education (Breen and Jonsson 2000, 2005; Müller and Karle 1993), or when analysing returns to education, for example how education determines individuals' income and socio-economic status

[1] GESIS – Leibniz Institute for the Social Sciences, Department Survey Design and Methodology, P.O. Box 12 21 55, 68072 Mannheim, Germany. Email: verena.ortmanns@gesis.org

Acknowledgments: I especially thank Silke Schneider, Michael Braun, Matthias Sand, Roberto Briceno-Rosas and Patricia Hadler for their helpful inputs and suggestions on this study as well as their feedback on earlier versions of this article. I also would like to thank the editor and especially the four anonymous reviewers for their remarks which helped to improve this article. I also thank my student assistants, Tim Dennenmoser, Svenja Friedel, Zeynep Umuc Demirci, Paweł Komendziński and Kimberly Herbst for their help with researching, collecting and coding the survey characteristics. I would also like to thank the UK Data Archive, Eurofound, the Norwegian Centre for Research Data, the GESIS Data Archive and the German Microdata Lab at GESIS for providing the data sets, the documentation of the survey characteristics and for their support. Finally, I thank Jane Roberts for proofreading. This work was supported by the Leibniz Association through the Leibniz Competition (formerly SAW Procedure) grant number SAW-2013-GESIS-5, as well as the European Union's Horizon 2020 research and innovation programme under grant agreement No 654221.

(Becker 1993; Blau and Duncan 1967; Bol and Van de Werfhorst 2013). Outside of stratification research, the education variable is an important proxy variable for another concept, such as cognitive competencies, and it is a widely used as a background or control variable. Quite often studies find a substantial effect of the education variable, for example when analysing values and behaviours, such as, political attitudes or voting behaviours (Bekhuis et al. 2014; Weakliem 2002), gender role attitudes (Bolzendahl and Myers 2004; Kalmijn 2003) or attitudes towards minorities and immigrants (Coenders and Scheepers 2003; Semyonov et al. 2008; Hyman and Wright 1979). In survey methodological research, the education variable is important because together with sex and age, it is often used to assess the comparability of the survey data, for instance with official data sources (Peytcheva and Groves 2009). Furthermore, education is often included when calculating post-stratification weights, which aim to correct for non-sampling errors such as nonresponse and may decrease the variance of a survey's estimate (e.g., ESS 2014b). Clearly the education variable is important for different purposes, and ideally should be of high measurement quality.

Previous research compared the education distribution across surveys within countries and years to assess how reliable the distribution of education is measured across surveys and thus how comparable the data are. For identical populations and time points, one would expect only minimal variation in the data. However, studies repeatedly revealed inconsistencies in education distributions across surveys even when they use the same harmonised education variables (Kieffer 2010; Ortmanns and Schneider 2016a, 2016b; Schneider 2009). These discrepancies indicate that the data cannot be comparable in some way. However, especially for cross-national comparative research, data need to be comparable. In more detail, the study of Kieffer (2010) observed discrepancies when comparing the distribution for the European Social Survey (ESS) with the EU-LFS for France. Large deviations were identified for the first three waves of the ESS in 2002, 2004 and 2006; while for 2008, the deviation was smaller. Schneider (2009), who compared data from 2002 to 2007, also identified inconsistencies when comparing the distributions for most countries in the European Union Statistics on Income and Living Conditions (EU-SILC), and in the ESS with the EU-LFS. Ortmanns and Schneider (2016b) replicated and extended this work by comparing education distributions for European countries included in four public opinion surveys between 2008 and 2012. They analysed the Eurobarometer, the European Values Study EVS, the International Social Survey Programme (ISSP) and the ESS, which was used as the reference survey. In the most comprehensive study to date, Ortmanns and Schneider (2016a) analysed seven cross-national survey programmes, again looking at the period 2008 to 2012. They included OECD's Programme for the International Assessment of Adult Competencies (PIAAC), EU-SILC, Eurobarometer, ESS, EVS and ISSP, and compared the education distributions for the same countries and years to the respective distribution in the EU-LFS. Since this study is the basis for this article, we will briefly summarise the main result to illustrate the problem. Ortmanns and Schneider (2016a) found that on average, 13% of respondents would have to change education categories to achieve an equal distribution with the EU-LFS. They also found substantial variation across surveys, ranging from 1% to almost 50%. These inconsistencies cannot reflect actual differences in the education distribution because it should be rather stable for the same country and year. Instead, these inconsistencies indicate a severe problem with data comparability across surveys, and thus methodological differences between the surveys must explain the observed deviations.

To date, researchers explain those inconsistencies commonly by differences in the measurement of education or the way country-specific response categories are classified into the International Standard Classification of Education (ISCED) (Kieffer 2010; Ortmanns and Schneider 2016a, 2016b). However, we cannot be sure that these are the only or most important factors just because they can be observed easily and are reported more often. Ortmanns and Schneider (2016a) identify single cases where they hypothesise that differences in the survey characteristics such as data collection modes, sampling designs, as well as selective unit nonresponse might also explain the inconsistencies because they do not find any problem in the measurement or the assignment of ISCED codes. Those survey characteristics refer to methodological aspects of a survey, and they differ across surveys because they are designed and organised differently, and apply different methodological standards. Thus, the survey characteristics influence the quality of the survey and its data. To systematically analyse and test the impact of surveys' methodological characteristics, we need an in-depth, quantitative and comprehensive analysis.

Such an analysis is conducted in this study, which analyses the impact of 15 survey characteristics and how they contribute to inconsistent education distributions across surveys within countries and years. As a starting point, we use the results from Ortmanns and Schneider (2016a), comparing the education distributions of six surveys with the EU-LFS for the years 2008 to 2012. We further extend the range of cross-national surveys by adding the Adult Education Survey (AES), the European Quality of Life Survey (EQLS), and the European Working Condition Survey (EWCS). Hence, this study compares the education distributions often cross-national surveys for 31 European countries. The research question is: Can survey characteristics explain the inconsistencies identified in the education distributions across surveys? Thirteen hypotheses are formulated and tested by estimating regression models.

Section 2 describes these cross-national surveys and how they measure education. It also introduces the challenges of comparing the education distributions and the survey characteristics across surveys. In Section 3, we present several different survey characteristics and derive our hypotheses regarding their contribution to the inconsistencies in education distributions. We use the Total Survey Error (TSE) framework (Groves et al. 2009; Groves and Lyberg 2010) to structure this presentation. In Section 4, the variables and methods are described, before presenting the results in Section 5. In Section 6, we discuss the results and limitations of the study and provide ideas for improving data comparability.

2. The Cross-National Surveys and their Education Measures

2.1. The Cross-National Surveys Covered in this Study

This study compares the education distributions of nine large-scale, cross-national surveys to the EU-LFS (Eurostat 2008, 2010a, 2011b, 2012), which we use as a benchmark, and estimates the impact of survey characteristics on the observed inconsistencies in the education distributions. To better understand the challenges of estimating the impact of survey characteristics when using the EU-LFS as a benchmark, and the consequences for the design of this study, we start with a brief description of the survey programmes.

Since the beginning of the EU-LFS in the 1970s, it has provided official household data for monitoring employment and unemployment in all EU countries and some European non-EU countries. The large number of countries included in the survey, the large sample sizes, the relatively high response rates and the probability-based sampling should produce representative high-quality data and thus an accurate education distribution for each country. Furthermore, the EU-LFS provides annual data, is fairly well documented, and it applies the official ISCED mappings. Thus, it is the most authoritative source regarding education data in Europe. Statistics based on the EU-LFS are, for instance, used in the annual OECD reports "Education at a Glance" (e.g., OECD 2015, 2016, 2017). EU-LFS data are also used when defining goals of the Europe 2020 strategy to enhance participation in education in all European countries (Eurostat 2019). The distribution of the EU-LFS education variable is also used as reference for other surveys, such as the ESS, when comparing or weighting data (ESS 2014a, 2014b). We are not aware of another official cross-national survey that fulfils all these criteria. Census data, for instance, typically do not provide harmonised data, which can be used for international comparisons; those have to be generated by the researcher herself. More important, to our knowledge, researchers cannot simply access an integrated data set of the latest official census data for all European countries. Hence, we use the EU-LFS as the benchmark survey in this study.

However, the EU-LFS also does not reflect the 'true' education distributions of the countries. The EU-LFS is an output harmonised survey, meaning the national surveys, to a large extent, are independent of each other and follow different national regulations. This applies for nearly all survey characteristics. Survey participation, for instance, is mandatory in roughly half of the countries the EU-LFS, but it is voluntary for the other countries. The response rate also varies greatly across countries between 30% and 98%. Furthermore, the countries use different sampling designs (simple or complex designs), as well as different modes of data collection (face-to-face, telephone, self-administered or mixed-mode). Of course, some guidelines and rules are specified to achieve as much comparable statistics as possible across countries, but the national survey designs entail quite different survey characteristics across the countries participating in the EU-LFS. This considerable variation in the survey characteristics of the EU-LFS forces us to analyse the impact of these survey characteristics with a rather broad approach. Therefore, we cannot assess which data collection mode causes more or fewer inconsistencies in the education distribution. Instead, we can only analyse whether mode differences between the survey in question and the EU-LFS affect the education distribution. As indicated, this applies to all survey characteristics; thus, we can only assess whether differences in the survey characteristics can contribute to inconsistencies in the education distributions across surveys within the same countries and years. This has to be considered when developing the hypotheses, and it adds complexity when operationalising the variables and interpreting the results. Nevertheless, it is important to mention that for all surveys, good documentation of the survey characteristics is an essential precondition for this study to identify how the survey characteristics differ across surveys within the same countries and years.

Another official survey included in this analysis is the EU-SILC (Eurostat 2010b). It was launched in 2003 with the aim of providing cross-sectional and longitudinal official micro-data on income, poverty, social exclusion, as well as living and housing conditions in the EU. We also analyse data from PIAAC (OECD 2013) and the AES (Eurostat 2011a), which focus on education. PIAAC is an OECD survey that measures adults' general basic skills, and first collected data in 2011/12 across OECD countries. The AES is a Eurostat survey that covers participation in formal and non-formal education and training of adults in EU countries. It began in 2007 and has been repeated nearly every fifth year. We also analyse data of the Eurobarometer (European Commission 2012), which was set up by the European Commission in the 1970s to monitor public attitudes towards the EU and related topics in all Member States. So far, the ISCED classification has only been implemented in three Eurobarometer studies, two of them have been conducted in 2010 and one in 2011. Additionally, we also analyse date from the EQLS (Eurofound 2014) and the EWCS (Eurofound 2011). Both surveys include all EU countries and they are funded through Eurostat and realised by Eurofound. The EQLS is conducted every four to five years since it was launched in 2003. The survey questions European citizens on general circumstances of their lives, such as employment, income, housing, family, happiness, and well-being. The EWCS was launched in 2005 and also runs quinquennially. It focuses on different aspects of employment, such as working time, learning and training, earnings and financial security, as well as work-life balance and health.

Lastly, three data sources from the academic community are included that cover different topics related to individuals' attitudes, beliefs, values and behaviour: the ESS (ESS 2016a, 2016b, 2016c), the EVS (EVS 2016), and the ISSP (ISSP Research Group 2015, 2016). The ESS was set up in 2002 and runs every second year in around 30 European countries. The EVS was launched in 1981, and data from five rounds of the survey are now available. The ISSP is an annual survey set up in 1985, and like PIAAC, it extends beyond Europe.

These surveys partly differ in the definition of their target population, for instance with regard to age groups. To render the samples as comparable as possible, we include only respondents aged 25 to 64 in all surveys. The EWCS focuses on people who are employed and thus, we restrict the analytic sample of the EU-LFS to employed respondents when comparing it to the EWCS.

2.2. Measuring and Comparing Educational Attainment in Cross-National Surveys

Asking respondents about their educational attainment is standard in almost all surveys in the social sciences. This question often refers to individuals' highest formal qualification or their highest completed educational level for which a diploma or certificate from a school, a formal vocational training or an institution of higher education or university is awarded. Respondents usually answer this question by selecting a category from a list. Those lists are necessarily country-specific, as education systems differ in their institutions and the names of the qualifications, which cannot be accurately translated (Braun and Mohler 2003; Schneider et al. 2016). Therefore, the ex-ante output harmonisation approach (Ehling 2003) is commonly used in cross-national surveys. Before data collection, the survey teams agree on a standard classification or a coding scheme and ideally set up guidelines specifying what has to be considered when developing the country-specific answer categories. The mapping of these categories to the standard classification, which is used to compare education across countries, is also developed in advance (Ehling 2003; OECD and Eurostat 2014). To harmonise the education categories across countries, most surveys choose the ISCED classification. This was designed by UNESCO in the 1970s and revised in 1997 and 2011. It aims to enable comparisons of country-specific education programmes for producing international education statistics. The ISCED classification defines international levels and types of education (UNESCO-UIS 2006), and education ministries and national statistical institutes map their educational programmes and qualifications to it. The most recent version of the classification was not yet implemented in most surveys for the years analysed, thus limiting this research to ISCED 97.

The main levels of ISCED 97 are:

- ISCED 0: Pre-primary education (or not completed primary education)
- ISCED 1: Primary education or first stage of basic education
- ISCED 2: Lower secondary or second stage of basic education
- ISCED 3: Upper secondary education
- ISCED 4: Post-secondary non-tertiary education
- ISCED 5: First stage of tertiary education
- ISCED 6: Second stage of tertiary education.

The focus here is on comparing the main levels of ISCED 97, ignoring the additional complementary dimensions on programme orientation, destination, duration and position in the national qualification structure, as most of the surveys analysed do not use them. All surveys we analysed implement the main levels of the ISCED classification or a variant thereof, from which we can derive the main level of ISCED 1997 for comparing the distributions. We need to aggregate ISCED levels 0 and 1 and levels 5 and 6 because those categories are not separated in all surveys. When comparing the EU-LFS and the ISSP, we also need to aggregate ISCED levels 3 and 4 (see Tables S1 and S2 in the online Supplemental material).

Following Ortmanns and Schneider (2016b, 2016a), we calculate Duncan's Dissimilarity Index (Duncan and Duncan 1955) to compare the education distributions between the EU-LFS, used as the benchmark survey, and the other surveys, which also use the ISCED classification. The index is defined as: $D = ½ \sum_{i=1}^{k} |x_i - y_i|$ where x_i denotes the number of observations in category i out of k ISCED categories for country A in year B in survey S, and y_i denotes the same for country A in year B in survey T. To interpret the resulting numbers as percentages, the index is rescaled to range from 0 to 100. This tells us how large the percentage is that needs to change categories to achieve equal education distributions between the EU-LFS and the survey in question.

Figure 1 shows the summary statistics of Duncan's Dissimilarity Index when comparing the education distributions between the EU-LFS and the other surveys within the same countries and years. The exact values can be found in Table S3 in the Supplemental material; these are used later as the dependent variable. We observe the smallest value of 1% in Duncan's index when comparing data for the Czech Republic from the 2010 EU-LFS and EU-SILC; this indicates nearly perfectly consistent data. The largest

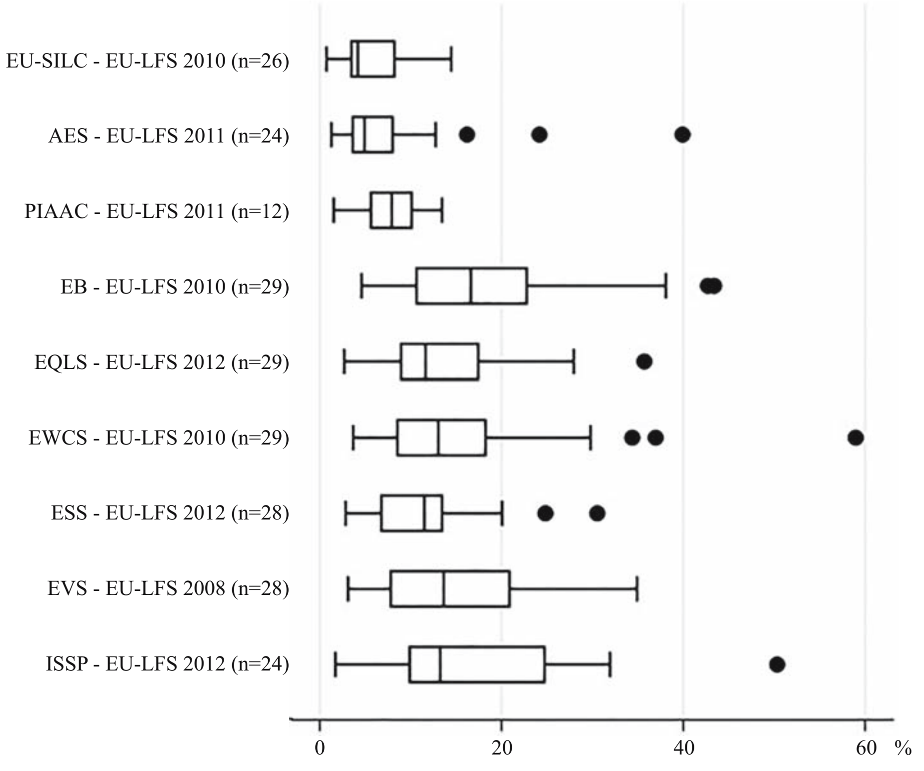

Fig. 1. Boxplots of Duncan's Dissimilarity Index across countries for all survey comparisons.
Notes: Here 'n' indicates the number of countries included in the analyses. Data sources, see in online Supplemental material.

deviation of 59% is found when comparing EU-LFS and EWCS data for Germany from 2010, which is even higher than the highest deviation identified by Ortmanns and Schneider (2016a). Overall, the mean inconsistency is almost 13%, meaning that on average 13% of respondents would need to change categories to achieve a distribution equal to that in the EU-LFS, which is the same result as found by Ortmanns and Schneider (2016a) based on a more limited set of international surveys. Duncan's Dissimilarity Index should, however, be close to zero because the education distributions should not vary across surveys when analysing the same country and year. This is clearly not the case. Looking at the individual surveys, we find the lowest discrepancy of roughly 6% when comparing the education distributions of the EU-LFS and the EU-SILC. When comparing the distributions of PIAAC and the AES to the EU-LFS, the discrepancy is 8%. We interpret these deviations as relatively small because they are clearly below the mean value of 13%. Duncan's index indicates a discrepancy of 12% between the ESS and the EU-LFS, 14% between the EQLS and the EU-LFS and 15% between the EVS and the EU-LFS. These percentages are around the mean value (between 10 and 15%) and, thus, we regard those as intermediate discrepancies. The comparison between the EWCS and the EU-LFS indicates a discrepancy of 16% and between the ISSP and the EU-LFS the discrepancy is 17%. We find the largest discrepancy of 19% when comparing the education distributions of the EU- LFS and the Eurobarometer. We interpret these deviations, which are above 15%, as larger inconsistencies.

3. Survey Characteristics

In order to explain differences between surveys, countries and years in terms of how well their education distribution matches that produced by the EU-LFS for the respective country and year, we refer to the Total Survey Error framework (Groves et al. 2009; Groves and Lyberg 2010) that describes different sources of errors that can appear at different stages of a survey. We use this framework for structuring the survey characteristics according to the different error sources, following the dimensions of representation of the population and measurement. An overview of all survey characteristics analysed in this study can be found in Table 1.

Considering that all surveys we analysed in this study are cross-national, we have to be aware that the survey characteristics do not only vary across surveys, but also across participating countries (Kohler 2008; Słomczyński et al. 2016). Different errors in the countries also reduce quality in terms of comparability across countries and/or surveys, as described in the application of the TSE approach to cross-national surveys (Smith 2010, 2011).

Some methodological survey characteristics are design features of the survey that can be changed in principle, such as the mode of data collection or fieldwork duration. Other survey characteristics, such as response rate, cannot be changed directly by the survey organisers. In methodological studies, the relationship between different kind for survey characteristics haven been examined as well as the impact of single characteristics on the data quality. For instance, studies have assessed whether the mode of data collection or offering incentives have an impact on response rates (Church 1993; Daikeler et al. 2019). Other studies evaluate the representation of the population of cross-national surveys by systematically comparing single survey characteristics across countries for a single survey (Kaminska and Lynn 2017) or across several surveys (Kohler 2007). Based on this research, best practice guidelines for survey organisers are developed (see e.g., Groves and Couper 1998, chap. 11).

3.1. Survey Characteristics Related to the Representation of the Population

In this section, we present several survey characteristics related to the representation of the population and how they could, theoretically, explain the inconsistencies in the education distributions between the EU-LFS, our benchmark, and the survey in question. When developing our hypotheses on the impact of the survey characteristics, we have to consider that those differ across countries also for the EU-LFS (see Subsection 2.1). Thus, we will only formulate undirected hypotheses indicating that differences in the survey characteristics of the EU-LFS and the survey in question might explain discrepancies in the education distributions across surveys within the same country and year.

Looking at the dimension of representation in the TSE approach, four kinds of errors are distinguished: coverage, sampling, unit nonresponse, and adjustment error (Groves et al. 2009). Coverage error emerges at an early stage even before drawing a sample; it arises when there is a discrepancy between the sampling frame and the target population.

Table 1. Overview of the survey characteristics and their operationalisation.

Dimension and errors of the TSE		Survey characteristic	Values	Values when comparing with EU-LFS
Representation of the population	Sampling error	Sampling design	Simple, complex	0 = equal, 1 = unequal
		Final sampling unit	Individual, household, dwelling/address	0 = equal, 1 = unequal
		Sample size	n	Absolute difference in the sample size divided by 1000
	Nonresponse error	Response rates	In percent	0 = equal response rate, 1 = higher, < 30 percentage points, 2 = lower, < 30 percentage points, 3 = lower, ≥ 30 percentage points, 4 = not available
		Survey participation	Mandatory, voluntary	0 = equal, 1 = unequal
		Fieldwork duration	Days	0 = equal duration, 1 = shorter, < 90 days, 2 = longer, > 90 days, 3 = longer, ≥ 90 days
	Sampling and nonresponse error	Index to validate probability sampling	Chance of interviewing a man/woman of a married couple living together in a two-person household	0 = equal, 1 = unequal
		Index on gender and age	Distribution of men and women for following age groups: 25–34, 35–44, 45–54, 55–64	Deviations in percent, indicating differences in the gender and age distribution

Table 1. Continued.

Dimension and errors of the TSE		Survey characteristic	Values	Values when comparing with EU-LFS
Measurement	Measurement error	Response categories of the education question		0 = same, 1 = similar, 2 = different
		Proxy-reporting	Yes, no	0 = equal, 1 = unequal
		Information taken from register	Yes, no	0 = equal, 1 = unequal
	Processing/ harmonisation error	Applying official ISCED mapping	Official ISCED mapping is applied, intended deviation, accidental deviation	0 = equal, 1 = unequal
		Degree of centralisation when applying ISCED	Decentralised, partly centralised, entirely centralised	0 = equal, 1 = unequal
Representation & measurement	Sampling, nonresponse, measurement and processing error	Mode of data collection	Face-to-face, telephone, self-administered, mixed-mode	0 = equal, 1 = unequal
		Fieldwork agency	Institute of public authority, university/scientific institute, commercial institute	0 = equal, 1 = unequal

Sampling error occurs when randomly taking a subset of sampling units from the sampling frame. When assessing sampling error, it is important to notice that most surveys analysed here use probability-based sampling methods, but that in the last stage, random-route approaches are applied in a few surveys. The survey characteristics on the sampling design and the final sampling unit reflect both coverage and sampling error and sample size only sampling error.

The *sampling design* influences the composition of the sample and thus also the education distribution. Almost every sampling design excludes some people from the target population, which might cause under- or over-coverage of certain groups (Groves and Couper 1998; Lohr 2009). In this article, we only distinguish between simple and complex sampling designs. In a simple design, the respondent is selected directly from an official register by means of a simple random sample. This is usually the case in the Scandinavian countries, which have central population registers. Ten countries of the EU-LFS have such a sampling design. In contrast, a complex sampling design might also use an official register, but multiple stages are used in the selection process. Other examples of a complex design are random digit dialing, and those where in the final stage a random route technique is applied. If the sampling design differs between the EU-LFS and the other survey, differences in the sample composition are likely, which might contribute to inconsistencies in the education distributions across surveys within the same countries and years (Hypothesis 1). Differences in the sample composition can also occur when both surveys apply complex sampling designs that differ from each other, for example through using different sampling frames. Unfortunately, generating a more detailed differentiation, for example by including additional information on the sampling frame, was not possible due to unstandardised or lacking information. For instance, it was also not possible to consider the information on how the surveys deal with institutionalised population because this often is not a central aspect in the documentation, although it is important to better assess errors in coverage and sampling (Schanze 2017).

Next, we look at the *final sampling unit*. We differentiate between an individual, a household or a dwelling/address. In most countries, the EU-LFS and the EU-SILC are household surveys and the dwelling/address or the household are the final sampling unit. Usually, in those surveys all respondents in a household above a specified age (15 in the EU-LFS, 16 in the EU-SILC), and more than one respondent at the same address or dwelling, are interviewed. This increases the chance of being selected to answer the questionnaire. In contrast, most other surveys use the individual respondent as the final sampling unit, and the individual probability of being selected is lower in these surveys (Groves et al. 2009). The different selection probabilities can influence the sample compositions and thus also the education distribution. To not overestimate the effect of the different sampling units, especially for the household surveys, data are weighted using available design weights. Therefore, we hypothesise that differences in the final sampling units across surveys might not affect the inconsistencies in the education distributions across surveys (Hypothesis 2).

The *sample size* of a survey matters because previous research shows that surveys with a larger sample size are more accurate, as the sampling error decreases (Biemer and Lyberg 2003). Surveys with smaller samples are more likely to have a sampling error that can lead to a slightly different sample composition and thus to a slightly different education distribution. All analysed surveys have rather large samples; however, the EU-LFS has by far the largest sample size for each country. Thus, we will definitely observe deviations in the sample size across the surveys. However, we estimate that these differences in the sample size might not contribute to the discrepancies in the education distribution (Hypothesis 3).

The nonresponse error, focusing on unit nonresponse, results in lacking representativeness of the sample. This error occurs if respondents systematically differ from non-respondents, that is sample members who refuse to participate in the survey or who cannot be interviewed. Here, we look at the following survey characteristics: mandatory survey participation, fieldwork duration and response rate. The survey characteristic on *mandatory survey participation* indicates that respondents are forced to participate in the survey. Usually, those surveys achieve higher response rates, and the nonresponse error is low because respondents who would refuse in voluntary surveys are often included in mandatory ones. Thus, we hypothesise that differences in mandatory survey participation across the EU-LFS and the other surveys might explain inconsistencies in the education distribution (Hypothesis 4). In the analysed surveys, participation is mandatory for only a small number of countries and surveys, namely 13 countries in the EU- LFS and nine in the AES.

Regarding fieldwork duration previous research indicate that longer field periods increase the chance of contacting and interviewing hard-to-reach respondents, whereas shorter fieldwork durations often leave less time for follow-ups. Thus, for surveys having a shorter fieldwork duration, errors of nonresponse become more likely (Biemer and Lyberg 2003). In the EU-LFS, fieldwork duration is usually three months and we distinguish whether the fieldwork compared to the EU-LFS is longer or shorter. We expect that different fieldwork durations—either considerably shorter or considerably longer than the benchmark—might increase inconsistencies in the education distribution across surveys within the same countries and years (Hypothesis 5).

The *response rate* is an important quality indicator and survey organisers invest a great deal of money in increasing it, for instance, by offering incentives to the respondents (Singer and Ye 2013; Groves et al. 2006). The response rate of the EU-LFS is relatively high, due to mandatory survey participation in some countries and because proxy-reporting is generally permitted. In contrast, for most other surveys the response rates are much lower and this might indicate that their realised samples can differ from the sample of the EU-LFS, that is, there is a higher risk of nonresponse error. Thus, we hypothesise that large differences in the response rates between the EU-LFS and the other surveys within countries and years could contribute to explaining inconsistencies in the education distributions (Hypothesis 6). However, we know that a high response rate alone is not enough to avoid nonresponse error (Bethlehem et al. 2011; Groves and Peytcheva 2008). Nevertheless, we decided to include this survey characteristic because we have no better indicator of the nonresponse bias.

The last error related to representation of the population is adjustment error. It emerges after data collection when calculating weights. This error is not taken into account in this study, because data are only weighted using design weights that correct for different inclusion probabilities due to different sampling designs across countries. Applying post-stratification weights that also correct for nonresponse errors is not feasible because those often correct for education, frequently by using the distribution of the EU-LFS as benchmark (e.g., ESS 2014b). This would lead to an (almost) equal distribution of the two surveys that are being compared.

Some specifications of the described survey characteristics relating to representation of the population are rather broad, for instance regarding the sampling design and sampling unit. This is caused by vague and sometimes also questionable documentation, particularly the design of the sampling process (for more information on the different standards in documentation, see Kohler 2008; Słomczyński et al. 2016). Therefore, it is advisable to also look directly into the data and check the realised representation. Firstly, we generate Sodeur's Index to validate probability sampling of the survey (Sodeur 1997, 2007). This index is based on the assumption that in a random sample, the chance of interviewing a man or a woman in a married couple living together in a two-person household is equal, namely 50:50. We adapt this and define the observed distribution of the EU-LFS as a benchmark. For calculation, we firstly restrict all samples to the 25 to 64 age group and married couples living in two-person households. Unfortunately but not unexpectedly, the required variables on marital status and household composition differ greatly across surveys, so adaptations are needed (for details see Annex 1 in Supplemental material). We calculate the gender distribution of this restricted sample and compare it to the distribution identified in the respective sample of EU-LFS, applying the following formula: $B_{UNR} = \frac{\hat{p}-p}{\sqrt{var(\hat{p})}}$ where p is the proportion of women in the EU-LFS and $\hat{p}$ is the proportion of women in the survey in question for the same country and year. Finally, the 95% confidence interval is calculated so we can decide whether the gender distribution between the EU-LFS and the other survey is equal or not within the same country and year. Secondly, we calculate an index to compare the gender and age distributions for four age groups (25–34, 35–44, 45–54, and 55–64) across surveys. Here, we again calculate Duncan's Dissimilarity Index (Duncan and Duncan 1955) and we use the distribution of the EU-LFS as benchmark.

3.2. Survey Characteristics Related to Measurement

On the measurement dimension of the TSE framework, there are three kinds of error that can occur: invalidity, measurement error, and processing error (Groves et al. 2009). Invalidity occurs when there is a disparity between the theoretical construct (what is intended be measured) and what is actually measured by the indicator. In this study, we do not expect to find invalidity because every survey asks respondents for their highest educational attainment in an equivalent way, asking respondents for their highest certificate/degree or their achieved educational level.

Measurement error occurs when a mismatch exists between the ideal measurement and the actual response obtained from the respondent. A potential source of measurement error across surveys is differences in the *response categories* in the education question. Previous research shows many examples pointing at differences in the measurement instrument as a source of inconsistent education data (Kieffer 2010; Ortmanns and Schneider 2016a, 2016b; Schneider 2009). For instance, when surveys use ambiguous terms or generic descriptions of educational qualifications, instead of the official name of the qualifications, the chance that the response categories differ across surveys is quite high. Thus, this survey characteristic seems to be of some importance when explaining inconsistencies in the education distributions. In the education question, the response categories are the key element influencing respondents' answers. All analysed surveys use country-specific response categories for the education question. To assess the similarity of the response categories of the EU-LFS and the other surveys, we qualitatively compared the education categories for every survey, country and year and generated an index. It distinguishes whether the categories are the same as, similar to, or different from the categories used in the EU-LFS. Detailed information on this index is provided in Annex 2 (Supplemental material). In general, we know that different stimuli can affect respondents' answers (Groves et al. 2009) and this also seems to occur with the education question, even though it is a factual question. Thus, different response categories are a probable explanation for inconsistencies in the education distributions (Hypothesis 7).

Relating to the measurement, we also measure whether *proxy-reporting* is allowed or prohibited. If the survey allows proxy-reporting, a respondent's partner or (adult) child might answer the questions instead of the selected respondent, or the 'head of the household' responds for every household member. Proxy-reporting can only be used in household surveys; thus, it applies to the EU-LFS, EU-SILC and the AES. Proxy-reporting is cognitively demanding, and measurement errors are likely due to lack of knowledge leading to incorrect answers (Blair et al. 2011; Kreuter et al. 2010; Moore 1988). Thus, we expect that differences in the allowability of proxy-reporting can contribute to inconsistencies in the education distribution across surveys (Hypothesis 8).

The last survey characteristic related to measurement error distinguishes *whether respondents' educational attainment is retrieved from a register* or not. Some countries, mostly Scandinavian ones, have population registers from which socio-demographic information, including education, can be directly retrieved. Register information is regarded as high quality and trustworthy (Biemer and Lyberg 2003). Therefore, differences in this survey characteristic on retrieving information from a register may explain inconsistencies in the education distribution (Hypothesis 9). However, we also have to be aware that register information is not free of errors either, due to delayed updates, especially for younger people who are currently in education (Kleven and Ringdal 2017). Only four countries of the EU-LFS use register information.

Next, we look at errors in the data processing, including harmonisation, these emerge while transforming responses into the final data set to be used for analysis. Processing errors seem to be of great importance: previous studies have repeatedly reported errors when classifying the country-specific educational qualifications into ISCED (Kieffer 2010; Ortmanns and Schneider 2016a; Schneider 2009; Hoffmeyer-Zlotnik 2008). Those errors directly influence the education distributions. We distinguish two survey characteristics here. The first one indicates *whether the official ISCED mapping is applied.* This is important because only if the educational qualifications are classified to ISCED in a consistent way, for example by following the official mappings, the education distributions are comparable across surveys (Schneider 2009). This characteristic distinguishes whether the assignment of ISCED codes to national education categories follows the official mapping or whether we find deviations from the official mapping. The EU-LFS and EU-SILC are conducted by the national statistical offices, which are also often responsible for developing countries' ISCED mapping, meaning they determine the ISCED code for each country-specific educational qualification. Therefore, we expect that the EU-LFS and the EU-SILC follow the official mapping and that processing errors are rare in these surveys. In the other surveys, classification errors may occur more often because of lack of expertise in implementing the ISCED classification, which might lead to 'accidental' errors. The other reason for this processing error is lack of trust in the official mappings and this might lead to intended deviations from the official ISCED mapping. This deviation is more common in academic surveys such as ESS, EVS and ISSP, which are not obliged to follow the official ISCED mappings. Therefore, we estimate that differences in the application of the official ISCED mappings across surveys can contribute to inconsistencies in the education distribution (Hypothesis 10).

The second survey characteristic indicating processing or harmonisation error describes the *degree of centralisation when applying the ISCED classification* for the survey. It distinguishes between decentralised, partly centralised and centralised processing. In the decentralised approach, the country teams, who are familiar with their education system, are responsible for assigning the ISCED codes to national education categories. The EU-LFS and most other surveys implemented this approach. In contrast, in the centralised approach, one institute is responsible for assigning the ISCED codes for all countries of the survey. The Eurobarometer follows this method. Applying ISCED codes for several countries requires much expertise in ISCED and in the different educational systems. If one of these components is lacking, the chance of processing or harmonisation errors increases. Another approach combines both methods: classifying the national education category in ISCED is carried out by the country teams, but it is also checked centrally. This is beneficial because it involves country experts and an expert in the application of ISCED, and aims to optimise cross-national comparability. The ESS implemented this approach. Hence, differences in the degrees of centralisation across the surveys can increase inconsistencies in the education distributions across surveys within the same countries and years (Hypothesis 11).

3.3. Survey Characteristics Related to Both Measurement and Representation

Two survey characteristics are related to both dimensions of the TSE framework: mode of data collection and fieldwork organisation. Regarding the *mode of data collection*, we distinguish between face-to-face interviews, telephone interviews, self-administered modes (including web and postal surveys), and mixed-mode designs. The mode is a relevant factor for representation because different modes tend to systematically over- or under-represent certain groups, for example web surveys tend to over-represent more highly educated respondents (Couper 2000; Dever et al. 2008). Regarding the measurement dimension, the mode indicates the presence of an interviewer and the communication channel used. In face-to-face or telephone interviews, the presence of an interviewer makes socially desirable answering and interviewer effects more likely (De Leeuw and Van der Zouwen 2001; Lyberg and Kasprzyk 2011), however, interviewers may also help the respondent identify a suitable answer. In face-to-face or self-administered modes, respondents usually see a list of education categories, while in telephone interviews, these categories are read out or an open response is coded by the interviewer, which is more error-prone and primacy or recency effects can occur in the former case (Noelle-Neumann and Petersen 2000). Therefore, we expect that different modes of data collection across the surveys within the same countries and years can increase inconsistencies in the education distributions across surveys (Hypothesis 12).

Fieldwork agencies are responsible for conducting the survey and are thereby involved in several aspects of sample representation and measurement. Therefore, the fieldwork agency can be seen as indicator for the standard of the survey and as proxy for different aspects, including those that could not be specified as survey characteristic due to a lack of information. This, for instance, applies to the availability of information on interviewer training. Concerning the EU-LFS, we would expect the overall standard to be quite high, largely because the fieldwork is done by a public authority, mostly the national statistical offices. This also applies to the second official survey, the EU-SILC. For the other surveys, commonly other fieldwork agencies are responsible, e.g., universities, other scientific or commercial institutes. We hypothesise that different kinds of fieldwork agencies can contribute to inconsistencies in the education distributions across surveys within the same countries and years (Hypothesis 13).

4. Data, Variables and Methods

In this study, we analyse the impact of surveys' methodological characteristics on discrepancies between the distributions of the harmonised education variable when comparing the EU-LFS with nine other surveys within the same countries and years. A description of the EU-LFS and the other surveys was already given in Subsection 2.1. This study focuses on these surveys from the period 2008 to 2012. If a survey was run several times during this time, such as the EU-SILC, the Eurobarometer, the ESS and the ISSP, it is only included once in order not to overestimate its effect. For most surveys the education distribution is stable over the years, as long as the country-specific measurement instruments and the harmonised education variable do not change (Ortmanns and Schneider 2016b, 2016a). When deciding which year to include, we consider the following factors: (a) number of countries covered, (b) completeness of documentation of survey characteristics, (c) whether its harmonised education variable has systematically changed (as in the ESS 2010 and the ISSP 2011), in which case the most recent year is included, (d) when a single country is not present in the selected year, information from an earlier round is used for this country. Due to a consequential processing error in the ISCED variable for Iceland in the EU-LFS 2011 and 2012 (for details see Ortmanns and Schneider 2016a), data before 2011 are included as far as possible. Thus, we include the EU-SILC and the Eurobarometer of 2010, and the ESS and ISSP of 2012.

As described in Subsection 2.2, the dependent variable is Duncan's Dissimilarity Index that compares the education distributions for each country and year of the EU-LFS with the respective country and year of each other survey. The independent variables reflect the survey characteristics (see Section 3) that differ across surveys for the same country-year comparison. Annex 3 (Supplemental material) provides basic descriptions of each survey characteristic. As mentioned, we focus on whether the survey characteristics differ between the EU-LFS and the respective other survey. Thus, most variables are coded as binary and distinguish whether the survey characteristics are 'equal' (0) or 'unequal' (1). The variables on response categories, fieldwork duration, response rates, sample size and the index of gender and age distribution are operationalised in a slightly more nuanced way. As described in Subsection 3.2, we generate an index to assess the comparability of the response categories and distinguish between equal, similar and different. When comparing the fieldwork duration of the EU-LFS with the other surveys, we distinguish between the following categories: ‘equal fieldwork duration to the EU-LFS’, including up to five percentage points more or fewer days than the EU-LFS, ‘longer duration: up to 90 days’ and ‘longer duration: 90 days or more’, ‘shorter duration: up to 90 days’. These four categories cover all comparisons. Regarding response rates, we use the ones reported in the survey documentation, even when we do not know exactly how these have been calculated, which may hamper their comparability. For the comparison of the response rates, we generate the following categories: ‘equal response rate to the EU-LFS’ if the response rate is up to 5 percentage points lower or higher than in the EU-LFS, ‘lower response rate: up to 30 percentage points, ‘lower response rate: 30 percentage points or more’ and ‘higher response rate: up to 30 percentage points’. A category indicating a higher response rate of more than 30 percentage points was not required. Unfortunately, the Eurobarometer does not provide information on response rates and for some countries of the other surveys the response rates are not documented. In order to be able to include those anyway, we generate an additional category ‘information not available’. The categories of the variables on fieldwork duration and response rate are based on their distributions, and in order to avoid small or empty categories, they are rather broad. We include these categories as dummy variables in the analysis, and the categories indicating equal response rate or fieldwork duration are used as reference categories. When comparing the sample sizes of the EU-LFS with the other surveys, we calculate the absolute differences in the sample size and then divide by 1,000 because of the very high number of respondents in the EU-LFS. We then include this as a continuous variable. Duncan’s index on the gender and age distribution delivers percentages and these are directly included in the regression models.

For many of the survey characteristics analysed, it would be desirable to use a higher level of detail. Unfortunately, this is not possible due to large variation in the accessibility of information, and especially the quality and the richness of the documentation. Still we had to exclude single countries in single surveys from the analysis when the information on a survey characteristic was not available. Thereby the data set is reduced from 248 to 229 survey comparisons and their respective comparisons of survey characteristics. The highest number of countries covered for one comparison is 29 when comparing EU-LFS with the Eurobarometer, or the EQLS or the EWCS, whereas the comparison between EU-LFS and PIAAC contains only 12 countries. An overview of the countries participating in the surveys and those included in the analysis can be found in Table S4 (Supplemental material).

Survey characteristics may correlate with each other and also with the survey programmes. Multicollinearity could make it hard to properly disentangle the effects of individual variables. Therefore, we checked the correlations between the different survey characteristics beforehand and Cramer’s V was below 0.65. More details can be found in the Tables showing cross tabulations and correlations for selected survey characteristics in Annex 4 (Supplemental material). Additionally, we calculate the Variance Inflation Factor (VIF) after each regression model.

In the analysis, we estimate four multiple OLS regression models to explore the impact of different survey characteristics on inconsistencies in the education distributions. The first model shows the impact of the survey programmes alone and thereby illustrates the large variation in the education distributions across surveys. The survey comparisons are included as dummy variables, and the comparison of EU-SILC and EU-LFS is used as reference. To explain these inconsistencies through differences in the survey characteristics, the second model adds the survey characteristics related to representation of the population. The third model includes survey characteristics related to measurement and survey programmes. To further reduce multicollinearity we calculate the final model excluding the dummy variables of the survey programmes. This model focuses on the survey characteristics that show statistically significant effects in Models 2 and 3.

5. Results

5.1. *Impact of the Survey Programmes*

As seen in the boxplot diagram (see Figure 1, Subsection 2.2) the inconsistencies in the education distributions differ strongly across surveys within the same countries and years. As expected, this pattern recurs when running a linear regression to predict Duncan's Dissimilarity Index by the survey programmes alone.

Model 1 in Table 2 shows low values for the regression coefficients for PIAAC ($b = 2.30$) and the AES ($b = 2.38$) and these survey comparisons are not statistically significant. The regression coefficients of the comparisons to the other survey programmes are higher ($b > 5.00$) indicating larger inconsistencies in the education distribution than in the reference comparisons of EU-LFS and EU-SILC. The comparison of the EU-LFS and the ESS is significant at the five percent level ($p < .05$), and the comparisons of the EU-LFS to the Eurobarometer, the EQLS, the EWCS, the EVS and the ISSP are highly significant ($p < .001$).

The adjusted R^2 of this model is 17%, meaning 17% of the variance can be explained by just the surveys themselves. This is unexpected because we can imagine the survey programmes as 'containers' for different survey characteristics. To identify which survey characteristics contribute to the inconsistencies in the education distributions, we estimate further regression models.

5.2. *Impact of Survey Characteristics Related to the Representation of the Population*

In addition to the first model, this model (Model 2 in Table 2) includes the survey characteristics related to the representation of the population, namely: sampling design, final sampling unit, sample size, mandatory survey participation, fieldwork duration, response rate, Sodeur's Index and Duncan's Dissimilarity Index for the age and gender distributions. Mode of data collection and fieldwork agency are also included.

This model shows that adding variables related to representation does not improve model fit: The adjusted R^2 of this model is also 17%. To estimate the quality of this model relative to the first model, we calculate the Akaike Information Criterion (AIC). For Model 1, the AIC is 1650.8 and for this model the AIC slightly increases to 1664.4. The model that shows the lowest value of the AIC, here Model 1, performs best. Regarding multicollinearity, the highest value of the VIF in this model is 7.1, which we observe for the dummy variable of the Eurobarometer. This indicates that the

Table 2. Results from regression analyses estimating the impact of survey characteristics on the inconsistencies in the education distribution across surveys.

	Model 1			Model 2			Model 3			Model 4		
Predictor	b	SE	p	b	SE	p	b	SE	p	b	SE	p
Survey:(ref: SILC)												
AES	2.38	2.47	.337	4.25	3.10	.172	−1.05	3.06	.731			
PIAAC	2.30	3.04	.451	4.42	3.72	.237	−3.08	4.21	.464			
EB	12.97***	2.36	<.001	14.94**	4.63	.001	3.71	7.19	.606			
EQLS	8.33***	2.36	<.001	11.70**	4.01	.004	−0.70	4.34	.872			
EWCS	10.21***	2.36	<.001	14.22**	4.06	.001	1.40	4.30	.745			
ESS	5.94*	2.38	.013	8.55*	3.62	.019	−2.74	6.78	.686			
EVS	9.20***	2.38	<.001	12.34**	3.69	.001	1.59	4.22	.708			
ISSP	11.57***	2.47	<.001	13.95***	3.88	<.001	0.50	4.36	.909			
Different sampling design (ref: equal)				3.67*	1.52	.016	3.37**	1.25	.007	3.43**	1.25	.007
Different sampling unit (ref: equal)				0.51	1.75	.772						
Differences in Sodeur's index (ref: equal)				−1.11	2.57	.665						
Duncan's index age/gender				−0.10	0.19	.594						
Sample size/1000				0.00	0.01	.647						
Fieldwork duration: (ref: equal)												
Shorter, < 90 days				1.09	2.28	.631						
Longer, < 90 days				0.16	2.22	.941						
Longer, ≥ 90 days				−0.73	2.53	.772						
Response rate: (ref: equal)												
Higher, < 30 percentage points				−0.90	3.11	.772						
Lower, < 30 percentage points				−3.55	2.66	.183						
Lower, ≥ 30 percentage points				−5.14	2.93	.081						
Not available				−4.01	3.91	.307						
Differences in mandatory participation (ref: equal)				0.82	1.32	.537						
Different mode (ref: equal)				2.15	1.43	.134	1.44	1.19	.225	1.46	1.19	.217
Different agency (ref: equal)				−1.42	2.42	.557	−0.31	2.17	.886	1.34	1.71	.432
Differences in proxy-reporting (ref: equal)							3.97	3.21	.217	2.89	1.98	.148
Difference in using register (ref: equal)							−0.65	1.93	.735	−0.47	1.92	.806

Table 2. Continued.

	Model 1			Model 2			Model 3			Model 4		
Predictor	b	SE	p	b	SE	p	b	SE	p	b	SE	p
Response categories: (ref: equal)												
Similar							0.87	2.39	.714	0.67	2.37	.776
Different							5.13*	2.33	.029	5.28*	2.24	.020
Differences in centralised coding (ref: equal)							0.53	5.67	.926	0.86	1.25	.494
Differences in ISCED coding (ref: equal)							9.20***	1.31	$<.001$	9.39***	1.27	$<.001$
Constant	5.58**	1.71	.001	5.63	3.62	.122	2.15	2.27	.346	1.84	2.18	.399
Adjusted R^2 (%)	16.61			16.67			35.59			34.00		
Akaike information criterion (AIC)	1650.76			1664.42			1600.04			1598.15		
Mean of variance inflation factor (VIF)	1.75			3.43			7.17			1.79		
Number of observations	229			229			229			229		

Notes: Data sources, see in online Supplemental material.

Eurobarometer correlates with the analysed survey characteristics. The mean value of the VIF of this model is 3.4, which is higher than in Model 1 (mean VIF of 1.8) but still unproblematic.

The only survey characteristic that has a statistically significant impact ($p < 0.05$) in this model is different sampling designs across the surveys. The regression coefficient of 3.7 indicates that different sampling designs increase the inconsistencies in the education distributions by roughly four percentage points, compared with equal designs. Thus, we do not reject hypothesis H1. From the results of this model, we find no evidence that the survey characteristics contribute to a higher inconsistency of the education distribution and therefore we do not reject H2 and H3 and we reject hypotheses H4 to H6, H12 and H13. In contrast to most survey characteristics, the survey effects remain significant and their regression coefficients even increase. Overall, this model shows that even when controlling for a substantial number of survey characteristics related to the representation of the population, the survey programmes themselves have by far the largest impact on the observed inconsistencies in the education distributions across surveys.

5.3. *Impact of Survey Characteristics Related to Measurement*

The third regression model shown in Table 2 focuses on the survey characteristics related to measurement. The following survey characteristics are included in this model: different response categories of the education question, proxy reporting, use of register information, applying of the official ISCED mappings and the degree of centralisation when applying ISCED. Also included are mode of data collection and fieldwork agency, which refer to both dimensions of the TSE, as well as the sampling design, which was significant in the second model. This model also controls for the survey programmes again.

This model has an adjusted R^2 of 36%, meaning more than one-third of the variance can now be explained. This is an increase of 19 percentage points compared to the previous models. The increase of the adjusted R^2 indicates a strong impact of survey characteristics related to measurement, over and above the effects of the surveys themselves. Compared to Models 1 and 2, the AIC decreases to 1600.0, which indicates a higher quality of this model. Concerning multicollinearity, the mean value of the VIF is 7.2, which is higher than in Models 1 and 2. In detail, we find high VIF values of around 20 for the dummy variables of the survey programmes for the Eurobarometer and the ESS, as well as the survey characteristic on the degree of centralisation when applying ISCED. This is not surprising because we know that this survey characteristic is strongly associated with the survey programme.

In this model, three survey characteristics have a statistically significant impact: different sampling designs, different response categories in the education item(s) and application of the official ISCED mapping. We find the strongest impact from the survey characteristic that indicates differences in whether the official ISCED mappings were applied between the EU-LFS and the surveys in question. This variable shows a high regression coefficient of 9.2, meaning inconsistency in the mapping of the national educational qualification into ISCED increases inconsistencies in the education distributions by roughly ten percentage points compared to consistent mapping. This effect is highly significant ($p < 0.001$). Thus, whether the official ISCED mappings are applied is a crucial factor that explains deviations in the education distributions across surveys within countries and years. Therefore, we do not reject Hypothesis H10.

The survey characteristic indicating different response categories in the education items between the EU-LFS and the other surveys is also significant ($p < 0.05$). The regression coefficient of 5.1 indicates that using different response categories raises inconsistencies in the education distribution across surveys by roughly five percentage points compared to equal response categories. Thus, we also do not reject Hypothesis H7.

The survey characteristic assessing different sampling designs between the EU-LFS and other surveys, which was the only significant factor in Model 2, is again significant. The regression coefficient increases to 3.4 and the p-value is smaller in this model ($p < 0.01$), thus we again do not reject Hypothesis H1 in this model. Nevertheless, the effect of sampling design is smaller compared to the coefficients related to measurement.

All other survey characteristics are not statistically significant. The survey comparisons themselves are also not significant any more. Thus, in this model we identified the survey characteristics causing inconsistencies in the education distributions across surveys, and we successfully opened 'the black box of the surveys'.

In the final model (Model 4 in Table 2) the adjusted R^2 slightly decreases to 34%. The AIC declines to 1598.2, which is lowest value across all models, indicating that this is the best model estimated. Though excluding the survey programmes, we also reduce multicollinearity and the mean value of the VIF decreases to 1.8. The statistical significance of the variables assessing different sampling designs ($p < 0.01$), different response categories ($p < 0.05$) and differences in the application of the official ISCED mapping ($p < 0.001$) between the EU-LFS and the other surveys remain. This highlights the importance of these three survey characteristics independently of the survey programmes. Thus, we do not reject Hypotheses H1, H2, H3, H7 and H10, but according to this analysis, we can reject all other hypotheses. This result emphasises a predominant effect of measurement, especially the consistency of applying the official ISCED mappings and consistent response categories in the education question. Those are the key elements when it comes to explaining the inconsistencies in the education distributions across surveys within countries and years.

6. Conclusion and Discussion

This article asked which survey characteristics could explain the inconsistencies in the education distributions when comparing nine cross-national surveys to the EU-LFS. To answer that question, the impact of 15 survey characteristics and the survey programmes themselves were estimated. The data set used for this analysis contains detailed macro-information concerning the survey characteristics for the countries and years of the ten surveys. The main finding of this study is that differences in applying the official ISCED mappings (H10), differences in the response categories of the education question across surveys (H7), as well as – but to a lesser degree – differences in the sampling designs of the surveys (H1), are systematically related to inconsistencies in the education distributions across surveys within the same countries and years. These results are in line with our expectation and also with previous research (Kieffer 2010; Schneider 2009; Ortmanns and Schneider 2016a, 2016b) that focused on the measurement of the education variable to explain inconsistent education distributions. Hence, the focus of previous studies was well justified. The comprehensive analysis of survey characteristics in this study additionally shows that apart from the sampling design, the survey characteristics related to the representation of the population do not cause inconsistencies in the education distribution across surveys.

To achieve higher consistency in the education distributions across surveys, survey organisers should, firstly, reduce the processing error by improving the assignment of the response categories of the education item to the ISCED classification. To make recommendations on how to reduce the processing error, we further need to distinguish whether the deviation from the official ISCED mapping occurs accidentally or whether it is intended. ‘Accidental’ errors, which are often caused by limited knowledge when assigning the national educational qualification to the ISCED classification, can be avoided through implementing additional quality checks and the application of the official ISCED mappings in principal (Ortmanns and Schneider 2016a).

In contrast, the intended deviations applied by some academic surveys aim to enhance comparability of cross-national education data across countries (Ortmanns and Schneider 2016a). This is justified because during the development and the implementation of the ISCED mappings it is vulnerable to political influence of education ministries and national statistical offices. The latter often develop the national ISCED mappings and they do not equally strictly apply the ISCED criteria. At the same time, some criteria formulated in the ISCED classification are rather vague and thus leave some room for interpretation. This explains why countries with similar qualification nevertheless classify them to different ISCED codes. The intended deviations made by academic surveys attempt to correct for this. However, these deviations also introduce incomparability across survey, notably with official surveys applying the official ISCED mappings, such as the EU-LFS and the EU-SILC. Intended deviations could be avoided when the quality control of the national ISCED mappings, for example through UNESCO, would become stricter. As this is currently not ensured, the international survey community has good reasons to find solutions to produce comparable education data for their own purpose. Academic surveys, for instance, could agree on applying an ‘alternative’ ISCED scheme that adjusts the official mappings to optimise comparability over time and space. This alternative version should be well-documented and contain recodes to the official mappings in order to still compare them with official education data.

The second important recommendation to achieve higher consistency in the education distributions across surveys is to improve the education item itself. We should aim for standardised country-specific education categories, which use a terminology that is equally understandable for everyone and avoid generic terms and descriptions. These categories can then be implemented in all surveys, national as well as international, that measure education as a background variable. Of course, no instrument will be without measurement error; however, if every survey uses the same instrument, the error will be consistent and this enhances data comparability. The development of these country-specific education categories and their assignment to ISCED should be done by a national expert group, which should consist of experts of the country-specific educational system, experts of ISCED and also representatives of the national statistical office, the education ministry as well as a survey expert. Ideally, also an expert in cross-national surveys should be included in the discussion to consider comparability in international surveys. Additionally, for countries having a similar educational system, for instance Germany, Austria and Switzerland or the UK and Ireland, it is also worthwhile to exchange their suggestions and, even better, to discuss shared issues. Then we can also better consider comparability across *countries*, which we did not look at in this article.

This study also faces some limitations. An obvious one is the small number of cases (n = 229), which might be problematic for testing such a large number of survey characteristics. However, focusing on whether the survey characteristics are equal or unequal across surveys prevents us from having small or even empty cells. The disadvantage of these variables is that they are quite generic, and it is not possible to, for instance, to identify which kind of fieldwork agency (public authority including statistical office, university or other scientific institute, commercial institute) causes more or less inconsistent education distributions. We can only tell whether differences in the fieldwork agencies between the survey in question and the EU-LFS affect deviations in the education distribution. This structure of the variables and the low case number furthermore do not allow calculation of more complex models or application of multilevel modeling.

Another limitation of this study is that it compares the education distribution using the 1997 version of ISCED, whereas surveys are increasingly implementing the more recent version – ISCED 2011. However, we are convinced that the current results would not be very different and we would still find inconsistencies when comparing the education distributions across surveys within countries and years. One change in ISCED 11 is a better differentiation of levels within tertiary education, so when surveys implement this new version, they will be paying particular attention to the codes of tertiary education. However, we observe the greatest inconsistencies for ISCED level 3 (upper secondary education), and also find deviations in the adjacent categories ISCED level 2 (lower secondary) and ISCED level 4 (post-secondary, non-tertiary). At these levels we find most of the ambiguous terms and generic descriptions used in the response categories of the surveys, especially with the vocational qualifications. These can also cause errors when assigning ISCED codes. The inconsistencies on these levels will not disappear when implementing ISCED 11, unless surveys start primarily to correct for accidental errors when assigning ISCED codes and update the country-specific response categories alongside the implementation of the new ISCED version. The ESS in 2010 undertook such a detailed check and updated its variables, and a similar review took place for the EVS 2017. The ISSP is currently considering how best to implement ISCED 11. The effort invested in the education variables in these surveys is likely to reduce inconsistencies in the education distribution in the future.

An output of this study is the data file of survey characteristics that is publicly available at the SowiDataNet|datorium (Ortmanns 2020). Until recently, survey characteristics have rarely been considered in substantive data analyses, and only few studies exist that include them (e.g., Heath et al. 2009; Van Tuyckom and Bracke 2014). The main reason that survey characteristics are often neglected is probably that collecting and harmonising this information requires considerable effort. Often the documentation of survey characteristics is neglected, meaning we have to look at several documents of varying quality, to be found on different webpages of the surveys or data archives. Sometimes we still cannot find complete information, and it is little standardised. More systematic and easily accessible documentation would be very helpful. This would enhance transparency and increase the possibility of developing standards on how to report survey characteristics. Some initiatives have begun by collecting, documenting and publishing information on methodological survey characteristics relevant for their specific projects. Such an initiative exists for official statistics within the online platform MISSY, which provides metadata of the EU-LFS and EU-SILC. A further initiative that recently has been completed is part of the EU project ‘Synergies for Europe’s Research Infrastructures in the Social Sciences’. In work package two, the sampling practices of European surveys have been documented to compare and finally improve them (Scherpenzeel et al. 2017). The ongoing research project on survey data harmonisation of the Polish Academy of Sciences in cooperation with Ohio State University also devotes substantial effort to documenting and harmonising data related to democratic values and protest behaviours (Słomczyński et al. 2018). Unfortunately, this study was already underway, so the outcomes of these initiatives could only be used for cross-checking. Finally, the IPUMS-International project, a collaboration of the University of Minnesota, National Statistical Offices, international data archives, as well as other international organisations, harmonises publicly available census data and provides a systematic inventory (Minnesota Population Center 2019). Unfortunately, it does not (yet) offer a harmonised ISCED variable that can be used for cross-national comparisons. However, all these projects will facilitate future studies like this, as well as substantive (rather than methodological) studies that would like to control for the impact of a single survey characteristic.

7. References

Becker, G.S. 1993. *Human Capital. A Theoretical and Empirical Analysis with Special Reference to Education*, 3rd ed. Chicago: The University of Chicago Press.

Bekhuis, H., Lubbers, M., Verkuyten. M. 2014. “How Education Moderates the Relation Between Globalization and Nationalist Attitudes.” *International Journal of Public Opinion Research* 26(4): 487–500. DOI: https://doi.org/10.1093/ijpor/edt037.

Bethlehem, J.D., F. Cobben, and B. Schouten. 2011. *Handbook of Nonresponse in Household Surveys*. Hoboken: John Wiley & Sons.

Biemer, P.P. and L.E. Lyberg. 2003. *Introduction to Survey Quality*. Hoboken: John Wiley & Sons.

Blair, J., G. Menon, and B. Bickart. 2011. “Measurement Effects in Self vs. Proxy Response to Survey Questions: An Information-Processing Perspective.” In *Measurement Errors in Surveys*, edited by P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Matthiowetz, S. Sudman, 145–166. Hoboken: John Wiley & Sons.

Blau, P.M. and O.D. Duncan. 1967. *The American Occupational Structure*. New York: John Wiley & Sons.

Bol, T. and H.G. Van de Werfhorst. 2013. “Educational Systems and the Trade-Off Between Labor Market Allocation and Equality of Educational Opportunity.” *Comparative Education Review* 57(2): 285–308.

Bolzendahl C.I. and D.J. Myers. 2004. “Feminist Attitudes and Support for Gender Equality: Opinion Change in Women and Men, 1974–1998.” *Social Forces* 83(2): 759–789. DOI: https://doi.org/10.1353/sof.2005.0005.

Braun, M. and P. Ph. Mohler. 2003. "Background Variables." In *Cross-Cultural Survey Methods*, edited by J.A. Harkness, F.J.R. Van de Vijver, and P. Ph. Mohler, 101–115. Hoboken: John Wiley & Sons.

Breen, R. and J.O. Jonsson. 2000. "Analyzing Educational Careers: A Multinomial Transition Model." *American Sociological Review* 65(5): 754–772. DOI: https://doi.org/10.2307/2657545.

Breen, R. and J.O. Jonsson. 2005. "Inequality of Opportunity in Comparative Perspective: Recent Research on Educational Attainment and Social Mobility." *Annual Review of Sociology* 31: 223–243. DOI: https://doi.org/10.1146/annurev.soc.31.041304.122232.

Church, A.H. 1993. "Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis." *Public Opinion Quarterly* 57(1): 62–79. DOI: https://doi.org/10.1086/269355.

Coenders, M. and P. Scheepers. 2003. "The Effect of Education on Nationalism and Ethnic Exclusionism: An International Comparison." *Political Psychology* 24(2): 313–343. DOI: https://doi.org/10.1111/0162-895X.00330.

Couper, M.P. 2000. "Web Surveys: A Review of Issues and Approaches." *Public Opinion Quarterly* 64(4): 464–494. DOI: https://doi.org/10.1086/318641.

Daikeler, J., M. Bosnjak, and K. Lozar-Manfreda. 2019. "Web Versus Other Survey Modes: An Updated and Extended Meta-Analysis Comparing Response Rates." *Journal of Survey Statistics and Methodology*. DOI: https://doi.org/10.1093/jssam/smz008.

De Leeuw, E.D. and J. Van der Zouwen. 2001. "Data Quality in Telephone and Face-to-Face Surveys: A Comparative Meta-Analysis." In *Telephone Survey Methodology*, edited by R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II, and J. Waksberg, 283–300. New York: John Wiley & Sons.

Dever, J.A., A. Rafferty, and R. Valliant. 2008. "Internet Surveys: Can Statistical Adjustments Eliminate Coverage Bias?" *Survey Research Methods* 2(2): 47–60. DOI: https://doi.org/10.18148/srm/2008.v2i2.128.

Duncan, O.D. and B. Duncan. 1955. "A Methodological Analysis of Segregation Indexes." *American Sociological Review* 20(2): 210–217. DOI: https://doi.org/10.2307/2088328.

Ehling, M. 2003. "Harmonising Data in Official Statistics: Development, Procedures, and Data Quality." In *Advances in Cross-National Comparison. A European Working Book of Demographic and Socio-Economic Variables*, edited by J.H.P. Hoffmeyer-Zlotnik and C. Wolf, 17–31. New York: Kluwer Academic/ Plenum Publishers.

ESS. 2014a. *Documentation of ESS Post-Stratification Weights*. Available at: https://www.europeansocialsurvey.org/docs/methodology/ESS1-5_post_stratification_weights_documentation.pdf (accessed December 2019).

ESS. 2014b. *Weighting European Social Survey Data*. Available at: https://www.europeansocialsurvey.org/docs/methodology/ESS_weighting_data_1.pdf (accessed December 2019).

ESS. 2016a. European Social Survey Round 4 Data 2008. Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC data file version 4.4.

ESS. 2016b. European Social Survey Round 5 Data 2010. Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC data file version 3.3.

ESS. 2016c. European Social Survey Round 6 Data 2012. Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC data file version 2.3.

Eurofound. 2011. *European Working Conditions Survey (EWCS) 2010*. Colchester: UK Data Archive. Study number 6971 data file version 1. DOI: https://doi.org/10.5255/UKDA-SN-6971-1.

Eurofound. 2014. *European Quality of Life Survey (EQLS) 2011–2012*. Colchester: UK Data Archive. Study number 7316 data file version 3. DOI: https://doi.org/10.5255/UKDA-SN-7316-2.

European Commission. 2012. Eurobarometer 73.2 & 73.3 (2-32010). GESIS Data Archive, Cologne. ZA5236 data file version 2.0.1. DOI: https://doi.org/10.4232/1.11473.

Eurostat. 2008. European Union Labour Force Survey (EU-LFS) 2008. User Database, data file version 2016.

Eurostat. 2010a. European Union Labour Force Survey (EU-LFS) 2010. User Database, data file version 2016.

Eurostat. 2010b. European Union Statistics on Income and Living Conditions (EU-SILC) 2010. User Database, data file version CROSS-2010-6.

Eurostat. 2011a. Adult Education Survey (AES). User Database, data file version 1.

Eurostat. 2011b. European Union Labour Force Survey (EU-LFS) 2011. User Database, data file version 2016.

Eurostat. 2012. European Union Labour Force Survey (EU-LFS) 2012. User Database, data file version 2016.

Eurostat. 2019. Smarter, Greener, More Inclusive? Indicators to Support The Europe 2020 Strategy. Luxembourg: Publication Office of the European Union. Available at: https://ec.europa.eu/eurostat/documents/3217494/10155585/KS-04-19-559-EN-N.pdf/b8528d01-4f4f-9c1e-4cd4-86c2328559de (accessed December 2019).

EVS. 2016. European Values Study 2008, 4th wave. Integrated Dataset. GESIS Data Archive, Cologne. ZA4800 data file version 4.0.0. DOI:https://doi.org/10.4232/1.12458.

Groves, R.M. and M.P. Couper. 1998. *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons.

Groves, R.M., M.P. Couper, S. Presser, E. Singer, R. Tourangeau, G.P. Acosta, and L. Nelson. 2006. "Experiments in Producing Nonresponse Bias." *Public Opinion Quarterly* 70(5): 720–736. DOI: https://doi.org/10.1093/poq/nfl036.

Groves, R.M., F.J. Fowler, M.P. Couper, J.M. Lepkowski, E. Singer, and R. Tourangeau. 2009. *Survey Methodology*, 2nd ed. Hoboken: John Wiley & Sons.

Groves, R.M. and L.E. Lyberg. 2010. "Total Survey Error: Past, Present, and Future." *Public Opinion Quarterly* 74(5): 849–879. DOI: https://doi.org/10.1093/poq/nfq065.

Groves, R.M. and E. Peytcheva. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis." *Public Opinion Quarterly* 72(2): 167–189. DOI: https://doi.org/10.1093/poq/nfn011.

Heath, A., J. Martin, and T. Spreckelsen. 2009. "Cross-National Comparability of Survey Attitude Measures." *International Journal of Public Opinion Research* 21(3): 293–315. DOI: https://doi.org/10.1093/ijpor/edp034.

Hoffmeyer-Zlotnik, J.H.P. 2008. "Harmonisation of Demographic and Socio-Economic Variables in Cross-National Survey Research." *Bulletin de Méthodologie Sociologique* 98: 5–24. DOI: https://doi.org/10.1177/075910630809800103.

Hyman, H.H. and C.R. Wright. 1979. *Education's Lasting Influence on Values*. Chicago: University of Chicago Press.

ISSP Research Group. 2015. International Social Survey Programme: Health and Health Care - ISSP 2011. GESIS Data Archive, Cologne. ZA5800 data file version 3.0.0. DOI: https://doi.org/10.4232/1.12252.

ISSP Research Group. 2016. International Social Survey Programme: Family and Changing Gender Roles IV - ISSP 2012. GESIS Data Archive, Cologne. ZA5900 data file version 4.0.0. DOI: https://doi.org/10.4232/1.12661.

Kalmijn, M. 2003. "Country Differences in Sex-Role Attittudes: Cultural and Economic Explanations." In *The Cultural Diversity of European Unity. Findings, Explanations and Reflections from the European Values Study*, edited by W. Arts, L.C.J.M. Halmann and J.A.P. Hagenaars: 311–337. Leiden: Koninklijke Brill.

Kaminska, O. and P. Lynn. 2017. "Survey-Based Cross-Country Comparisons Where Countries Vary in Sample Design: Issues and Solutions." *Journal of Official Statistics* 33(1): 123– 136. DOI: https://doi.org/10.1515/JOS-2017-0007.

Kieffer, A. 2010. "Measuring and Comparing Levels of Education: Methodological Problems in the Classification of Educational Levels in the European Social Surveys and the French Labor Force Surveys." *Bulletin de Méthodologie Sociologique* 107: 49–73. DOI: https://doi.org/10.1177/0759106310369974.

Kleven, Ø. and K. Ringdal. 2017. Level of Education – Measuring the Quality of Questions in Survey Interviews by Administrative Records on Education. Experiences from the Norwegian European Social Survey 2004 – 2014. In *Proceedings of the 7th Conference of the European Survey Research Association (ESRA)*, July 20, 2017. Lisbon. Available at: https://www.europeansurveyresearch.org/conference/programme2017?sess=210#582 (accessed April 2020).

Kohler, U. 2007. "Surveys from Inside: An Assessment of Unit Nonresponse Bias With Internal Criteria." *Survey Research Methods* 1(2): 55–67. DOI: https://doi.org/10.18148/srm/2007.v1i2.75.

Kohler, U. 2008. "Assessing the Quality of European Surveys - Towards an Open Method of Coordination for Survey Data." In *Handbook of Quality of Life in the Enlarged European Union*, edited by J. Albers, T. Fahey, and C. Saraceno, 405–423. London, New York: Routledge.

Kreuter, F., S. Eckman, K. Maaz, and R. Watermann. 2010. "Children's Reports of Parents' Education Level: Does It Matter Whom You Ask and What You Ask About?" *Survey Research Methods* 4(3): 127–138. DOI: https://doi.org/10.18148/srm/2010.v4i3.4283.

Lohr, S.L. 2009. *Sampling: Design and Analysis*. 2nd ed. Boston: Brooks/Cole Cengage Learning.

Lyberg, L.E. and D. Kasprzyk. 2011. "Data Collection Methods and Mesurement Error: An Overview." In *Measurement Errors in Surveys*, edited by P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowwetz, and S. Sudman, 235–257. Hoboken: John Wiley & Sons.

Minnesota Population Center. 2019. *Integrated Public Use Microdata Series, International*. Data Minneapolis, MN: IPUMS, data file version 7.2. DOI: https://doi.org/10.18128/D020.V7.2.

Moore, J.C. 1988. "Self/Proxy Response Status and Survey Response Quality: A Review of the Literature." *Journal of Official Statistics* 4(2): 155–172.

Müller, W. and W. Karle. 1993. "Social Selection in Educational Systems in Europe." *European Sociological Review* 9(1): 1–23. DOI: https://doi.org/10.1093/oxfordjournals.esr.a036652.

Noelle-Neumann, E. and T. Petersen. 2000. "Das halbe Instrument, Die halbe Reaktion. Zum Vergleich von Telefon- und Face-to-Face Umfragen." In *Methoden in Telefonumfragen*, edited by V. Hüfken, 183–200. Wiesbaden: Westdeutscher Verlag.

OECD. 2013. Programme for the International Assessment of Adult Competencies (PIAAC) 2011 - Public Use Files: data file version 1.0. Available at: https://www.oecd.org/skills/piaac/data/ (accessed April 2020).

OECD. 2015. *Education at a Glance 2015: OECD Indicators*. OECD Publishing. Available at: https://www.oecd-ilibrary.org/education/education-at-a-glance-2015_eag-2015-en (accessed December 2019).

OECD. 2016. *Education at a Glance 2016: OECD Indicators*. OECD Publishing. Available at: https://www.oecd-ilibrary.org/education/education-at-a-glance-2016_eag-2016-en (accessed December 2019).

OECD. 2017. *Education at a Glance 2017: OECD Indicators*. OECD Publishing. Available at: https://www.oecd-ilibrary.org/education/education-at-a-glance-2017_eag-2017-en (accessed December 2019).

OECD, and Eurostat. 2014. *Joint Eurostat-OECD Guidelines on the Measurement of Educational Attainment in Household Surveys*. Available at: http://ec.europa.eu/eurostat/documents/1978984/6037342/Guidelines-on-EA-final.pdf (accessed December 2019).

Ortmanns, V. and S.L. Schneider. 2016a. "Can We Assess Survey Representativeness of Cross-National Surveys Using the Education Variable?". *Survey Research Methods* 10(3): 189– 210. DOI: https://doi.org/10.18148/srm/2016.v10i3.6608.

Ortmanns, V. and S.L. Schneider. 2016b. "Harmonization Still Failing? Inconsistency of Education Variables in Cross-National Public Opinion Surveys." *International Journal of Public Opinion Research* 28(4): 562–582. DOI: https://doi.org/10.1093/ijpor/edv025.

Ortmanns. V. 2020. Education distributions and survey characteristics of ten cross-national surveys. GESIS, SowiDataNet|datorium. Data file Version: 1.0.0. DOI: https://doi.org/10.7802/1.2002.

Peytcheva, E. and R.M. Groves. 2009. "Using Variation in Response Rates of Demographic Subgroups as Evidence of Nonresponse Bias in Survey Estimates." *Journal of Official Statistics* 25(2): 193–201.

Schanze, J.-L. 2017. *Report on Sampling Practices for the Institutionalized Population in Social Surveys*. Deliverable 2.16 of the SERISS project funded under the European Union's Horizon 2020 research and innovation programme GA No: 654221. Available at: www.seriss.eu/resources/deliverables (accessed December 2019).

Scherpenzeel, A., A.M. Maineri, J. Bristle, S. Pflüger, I. Mindarova, S. Butt, S. Zins, T. Emery, and R. Luijkx. 2017. *Report on the Use of Sampling Frames in European Studies*. Deliverable 2.1 of the SERISS project funded under the European Union's Horizon 2020 research and innovation programme GA No: 654221. Available at: www.seriss.eu/resources/deliverables (accessed December 2019).

Schneider, S.L. 2009. *Confusing Credentials: The Cross-Nationally Comparable Measurement of Educational Attainment*. (PhD thesis). Oxford: University of Oxford, Nuffield College. Available at: http://ora.ouls.ox.ac.uk/objects/uuid:15c39d54-f896-425b-aaa8-93ba5bf03529 (accessed December 2019).

Schneider, S.L., D. Joye, and C. Wolf. 2016. "When Translation Is Not Enough: Background Variables in Comparative Surveys." In T*he SAGE Handbook of Survey Methodology*, edited by C. Wolf, D. Joye, T.W. Smith, and Y.-c. Fu, 288–307. London: Sage Publications.

Semyonov, M., R. Raijman, and A. Gorodzeisky. 2008. "Foreigners' Impact on European Societies Public Views and Perceptions in a Cross-National Comparative Perspective." *International Journal of Comparative Sociology* 49(1): 5–29. DOI: https://doi.org/10.1177/0020715207088585.

Singer, E. and Ye, C. 2013. "The Use and Effects of Incentives in Surveys." *The ANNALS of the American Academy of Political and Social Science* 645(1): 112–141. DOI: https://doi.org/10.1177/0002716212458082.

Słomczyński, K.M., Jenkins, J.C., Tomescu-Dubrow, I., Kołczyńska, M., Wysmułek, I., Oleksiyenko, O., Powałko, P. and Zielinski, M.W. 2018. "Survey Data Recycling Dataverse - Master Box." Harvard Dataverse, data file version 1.1. DOI: https://doi.org/10.7910/DVN/VWGF5Q.

Słomczyński, K.M., Tomescu-Dubrow, I., Jenkins, J.C., Kołczyńska, M., Powałko, P., Wysmułek, I., Oleksiyenko, O., Zielinski, M.W. and Dubrow, J.K. 2016. *Democratic Values and Potest Behavior - Harmonization of Data from International Survey Projects*. Warsaw: IFiS Publishers.

Smith, T.W. 1995. "Some Aspects of Measuring Education." *Social Science Research* 24(3): 215–242. DOI: https://doi.org/10.1006/ssre.1995.1008.

Smith, T.W. 2010. "The Globalization of Survey Research." In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, edited by J.A. Harkness, M. Braun, B. Edwards, T.P. Johnson, L.E. Lyberg, P.Ph. Mohler, B.-E. Pennell, and T.W. Smith, 477–484. Hoboken: John Wiley & Sons.

Smith, T.W. 2011. "Refining the Total Survey Error Perspective." *International Journal of Public Opinion Research* 23(4): 464–484. DOI: https://doi.org/10.1093/ijpor/edq052.

Sodeur, W. 1997. "Interne Kriterien zur Beurteilung von Wahrscheinlichkeitsauswahlen." In *ZA-Information / Zentralarchiv Für Empirische Sozialforschung*, 41: 58–82. Cologne. Available at: https://www.ssoar.info/ssoar/handle/document/19999 (accessed December 2019).

Sodeur, W. 2007. “Entscheidungsspielräume von Interviewern bei der Wahrscheinlichkeitsauswahl - Ein Vergleich von ALLBUS-Erhebungen.” *MDA - Methoden Daten Analysen* 1(2): 107– 130.

UNESCO-UIS. 2006. *International Standard Classification of Education ISCED 1997*. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000146967?posInSet=1&queryId=a8d996ff-eddd-4153-9468-5d7f9e8f7199 (accessed December 2019).

Van Tuyckom, C. and P.F. Bracke. 2014. “Survey Quality and Cross-National Sports Research: A Case Study of the 2007 ISSP Survey.” *European Journal of Sport Science* 14(1): 228– 234. DOI: https://doi.org/10.1080/17461391.2012.683814.

Weakliem, D.L. 2002. “The Effects of Education on Political Opinions: An International Study.” *International Journal of Public Opinion Research* 13(2): 141–157. DOI: https://doi.org/10.1093/ijpor/14.2.141.

Received August 2018
Revised February 2019
Accepted December 2019